Modelling and Optimization for Green Transition of Complex Energy Networks

Prof. Luigi Glielmo

The Department of Electrical Engineering and Information Technologies
University of Naples Federico II

October 26, 2024

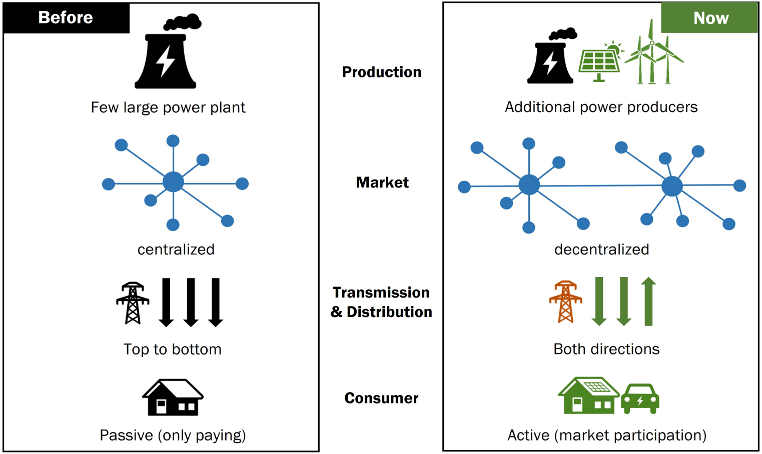
CCCN 2024
The 20th National Conference on Complex Networks

Overview

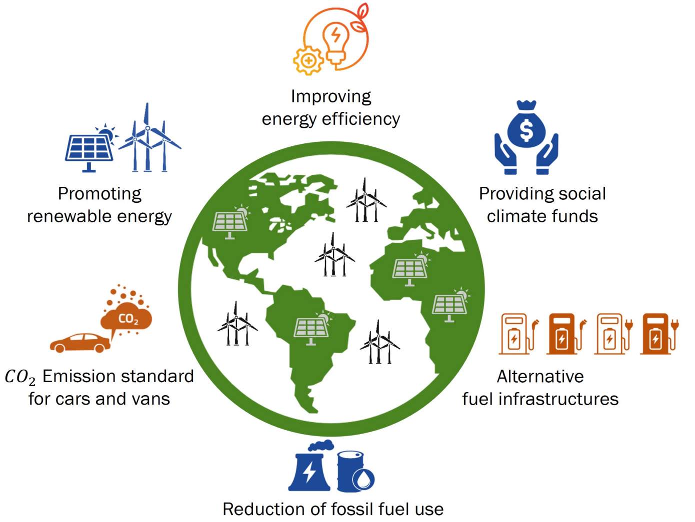
1. Evolution and challenges of energy networks
2. The European Initiatives
3. Smart Grids, Microgrids, Renewable Energy Communities
4. (A Part of) Our work on energy
5. Conclusions

Evolution and challenges of energy networks

Evolution of energy networks



Initiatives towards green transition



Initiatives towards energy transition

European Union (EU) goals

- **European green deal:** Aim for climate neutrality by 2050
- **Fit for 55 package:** 55% emissions reduction target by 2030
- **REPowerEU plan:** reduce dependence on fossil fuels

EU's Renewable Energy Directive (RED II) goals

- Promote renewable energy use across power, heating/cooling, and transport sectors
- Aim at 45% renewable energy share by 2030 in the EU's energy mix
- Support for renewable energy communities (RECs) and self-consumption models

Initiatives towards energy transition

European Union (EU) goals

- **European green deal:** Aim for climate neutrality by 2050
- **Fit for 55 package:** 55% emissions reduction target by 2030
- **REPowerEU plan:** reduce dependence on fossil fuels

EU's Renewable Energy Directive (RED II) goals

- Promote renewable energy use across power, heating/cooling, and transport sectors
- Aim at 45% renewable energy share by 2030 in the EU's energy mix
- Support for renewable energy communities (RECs) and self-consumption models

Smart grids

Actors

The smart grid includes a range of intermediate entities called as “actors”, such as:

- **Generator**: companies that generate electricity and sell in the wholesale market.
- **Energy retailer**: companies who purchase from the generators and sell to the end-users.
- **Aggregator**: a decision making body for a cluster of end-users.

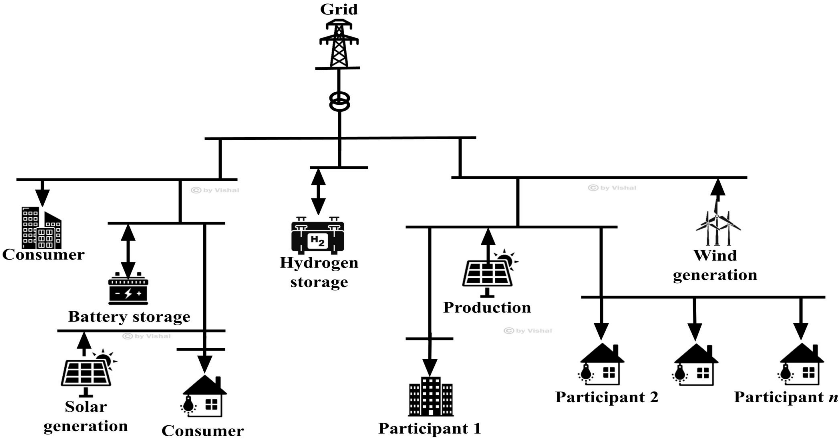
Novel concepts

The interaction among these actors at different levels in the energy network hierarchy has resulted in concepts such as

- **Microgrid/Nanogrid**
- **Renewable Energy Communities**

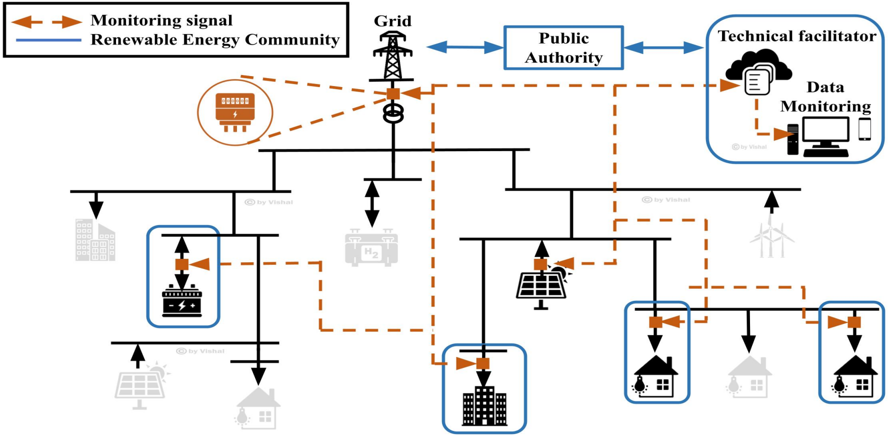
Microgrid

A microgrid is a group of interconnected loads and distributed energy resources that acts as a **single controllable entity** with respect to the grid.



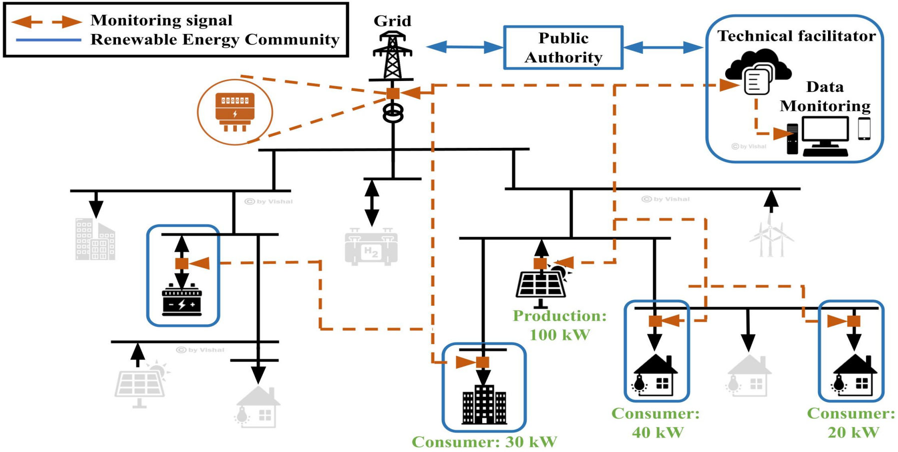
Renewable energy communities (RECs)

A REC is a group of producers and consumers who **legally agree** to generate, share, and manage local renewable energy to promote sustainability.



REC: notion of "virtual consumption"

- An example of a REC with 1 prosumer and 2 consumers (virtual power consumption of 90 kW):



Our work on energy

Research summary

Ongoing research activities

- Sizing and Planning of **Renewable Energy Communities**
- **Flexibility coordination / Demand response for microgrids**
- **Lithium-ion batteries** modeling and management
- **Hydrogen storage systems** modeling and management

Tools

- Optimization
- Optimal and model predictive control, nonlinear systems
- Data analysis and **(Multi-Agent) Reinforcement Learning**

Research summary

Ongoing research activities

- Sizing and Planning of **Renewable Energy Communities**
- **Flexibility coordination / Demand response for microgrids**
- **Lithium-ion batteries** modeling and management
- **Hydrogen storage systems** modeling and management

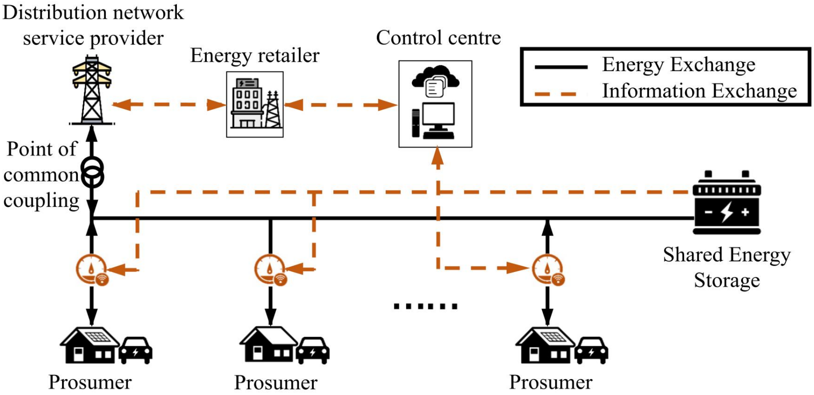
Tools

- Optimization
- Optimal and model predictive control, nonlinear systems
- Data analysis and **(Multi-Agent) Reinforcement Learning**

“Flexible” Microgrids

Residential microgrid

- Minimize individual electricity bill while coordinating the individual/shared flexibility.



Challenges and Alternatives

Challenges of Model-Based Approaches

1. a prerequisite to solving model-based problems is the knowledge of the future exogenous components, such as the demand, renewable generation and the like.
2. iterative nature of the distributed approaches.

An alternative?

To use the available "historical" data and learn control policies such that the agents can control the flexible load in a purely-decentralized manner.

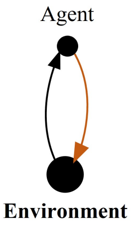
Learning from Historical Data

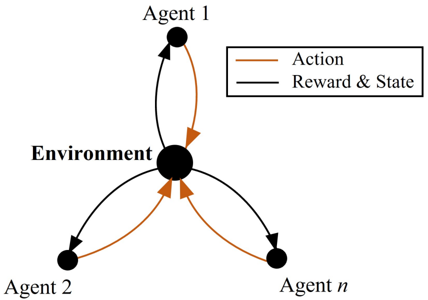
Reinforcement Learning (RL)

. . . where an agent learns a control policy fulfilling specific objectives, through interactions with its own environment. . .

Multi Agent RL

. . . when multiple agents share a common environment with the objective of learning individual control strategy. . .





Markov Decision Process

The interaction of a single agent with its environment can be modelled as a Markov Decision Process (MDP). A MDP can be represented by the following components:

- environment's state space $\mathcal{S}$
- agent's action space $\mathcal{A}$
- state transition probability function $f : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$
- reward function $\rho : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$
- discount factor $\gamma \in [0, 1)$

Assuming the agent follows a stationary policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$, the discounted sum of future rewards over the horizon (also known as return) is given by $G = \sum_{i=0}^{\infty} \gamma^i \rho^i$.

("Model-Free") Q-learning for A Single Agent

The objective therefore is to learn a policy π^*, which maximizes the expected return. To do so, given policy $\pi \in \Pi$, a value is associated with each state, called as the **value function** $V_\pi : \mathcal{S} \to \mathbb{R}$, given as

$$V_\pi(s) = \mathbb{E}_\pi[G|s], \quad \forall s \in \mathcal{S}, \quad s \text{ initial state.} \tag{1}$$

Likewise, a value is associated with each initial state conditioned on a given initial action, called as the **action-value function** $Q_\pi(s,a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, given as

$$Q_\pi(s,a) = \mathbb{E}_\pi[\rho' + \gamma V_\pi(s')], \quad \forall s, s' \in \mathcal{S}, a \in \mathcal{A}. \tag{2}$$

Watkins C., and P. Dayan. "Q-learning", *Machine learning* (1992)

contd. . .

Notion of optimality

- An optimal policy π^* is the one which maximizes the value function for all the initial states, i.e., $\pi^* = \arg\max_{\pi \in \Pi} V_\pi(s), \forall s \in \mathcal{S}$.
- All optimal policies share the same optimal value function, i.e., $V^*(s) := V_{\pi^*}(s)$.
- All optimal policies share the same optimal action-value function, i.e., $Q^*(s,a) := Q_{\pi^*}(s,a)$.

Recursive update rule in Q-learning

$$Q(s,\ a) \leftarrow (1-\alpha)Q(s,\ a) + \alpha\left[\rho' + \gamma \max_{a' \in \mathcal{A}} Q(s', a')\right]$$

The Q-function (a digression)

Consider Bellman's basic **backward iteration** in **deterministic** setting

$$V_k(s) = \max_a \left[\rho(s, a, s') + \gamma\, V_{k+1}(s')\right]$$

and let

$$Q_k(s, a) = \rho(s, a, s') + \gamma\, V_{k+1}(s').$$

It is straightforward to see that the backward iteration can be entirely written in terms of the Q-function as follows

$$Q_k(s, a) = \rho(s, a, s') + \gamma \max_{a'} Q_{k+1}(s', a').$$

Generalizing to a **stochastic** setting and **swapping** k **and** $k+1$, we get a **forward iteration**

$$Q_{k+1}(s, a) = \mathbb{E}\Big[\rho(s, a, s') + \gamma \max_{a'} Q_k(s', a')\Big].$$

The Q-function (a digression)

Consider Bellman's basic **backward iteration** in **deterministic** setting

$$V_k(s) = \max_a \left[\rho(s, a, s') + \gamma\, V_{k+1}(s') \right]$$

and let

$$Q_k(s, a) = \rho(s, a, s') + \gamma\, V_{k+1}(s').$$

It is straightforward to see that the backward iteration can be entirely written in terms of the Q-function as follows

$$Q_k(s, a) = \rho(s, a, s') + \gamma\, \max_{a'} Q_{k+1}(s', a').$$

Generalizing to a **stochastic** setting and **swapping** k **and** $k+1$, we get a **forward iteration**

$$Q_{k+1}(s, a) = \mathbb{E}\Big[\rho(s, a, s') + \gamma\, \max_{a'} Q_k(s', a') \Big].$$

Markov Game

The interaction of more than one agent with the environment and their interaction with each other can be modelled as a Markov Game (MG). A MG can be represented by the following components:

- set of agents $\mathcal{N} := \{1, \ldots, n, \ldots, N\}$
- environment's state space $\boldsymbol{\mathcal{S}}$
- joint action set $\boldsymbol{\mathcal{A}} := \mathcal{A}_1 \times \cdots \times \mathcal{A}_n \times \cdots \times \mathcal{A}_N$
- state transition probability function $f : \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{A}} \times \boldsymbol{\mathcal{S}} \rightarrow [0, 1]$
- n-th agent reward function $\rho_n : \boldsymbol{\mathcal{S}} \times \boldsymbol{\mathcal{A}} \times \boldsymbol{\mathcal{S}} \rightarrow \mathbb{R}$
- discount factor $\gamma \in [0, 1)$

Littman M. L., "Markov games as a framework for multi-agent reinforcement learning", *Machine Learning Proceedings* (1994).

Q-learning for multiple agents

Hyper-Q function

Let $\boldsymbol{\pi} := \{\pi_1, \ldots, \pi_n, \ldots, \pi_N\}$ be the joint control policy and accordingly

$$Q_n^{\boldsymbol{\pi}}(s, a_n,\ a_{-n}) := \mathbb{E}_{\boldsymbol{\pi}}\left[G_n \,|\, s, a_n, a_{-n}\right]. \tag{3}$$

In other words, the expected return achieved by agent n when, in state s, it takes a control action a_n, the remaining agents take a joint control action a_{-n}, and thereafter they follow the joint policy $\boldsymbol{\pi}$.

Recursive update rule . . .

$$Q_n(s,\ a_n,\ a_{-n}) \leftarrow (1-\alpha_n)Q_n(s,\ a_n,\ a_{-n})) + \alpha_n\left[\rho_n' + \gamma \max_{a_n' \in \mathcal{A}_n} Q_n(s', a_n', a_{-n}'))\right] \tag{4}$$

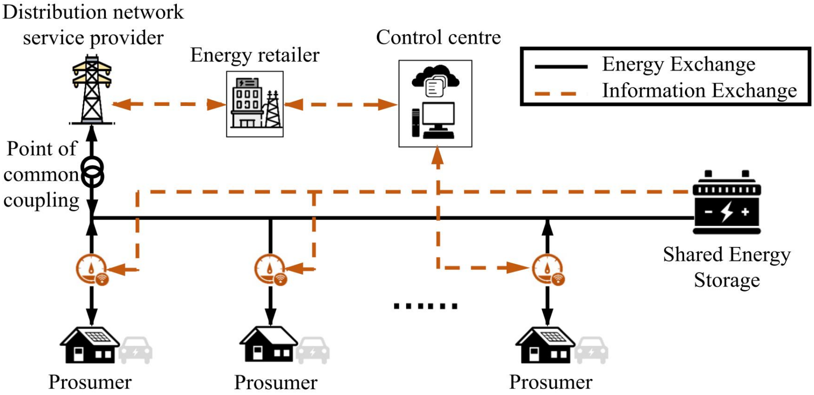
Tesauro G., "Extending Q-learning to general adaptive multi-agent systems", *Advances in Neural Information Processing Systems* (2003).

Maintaining Scalability

Each agent must rely on a limited "observation" of the state of the environment, i.e., the other agents' states

A Case of “Shared” Storage

Shared Energy Storage (SES)



SES modelling

The agents act "rationally" and can decide on how much power to exchange with the SES and the grid. All the power exchanges are associated with an economic exchange. We model that

- **the agents are billed/compensated if they discharge/charge the SES,**
- **the agents are billed/compensated if they withdraw/inject energy from/into the grid.**

The cost associated with different power exchanges (a.k.a. the electricity bill) can be expressed as

$$b_n^t = \max\{g_n^t c_{\mathrm{ToU}}^t,\, g_n^t c_{\mathrm{FiT}}^t\} - \min\{u_n^t c_{\mathrm{SES}}^{\mathrm{ch},t},\, u_n^t c_{\mathrm{SES}}^{\mathrm{dch},t}\}. \tag{5}$$

In (5), $g_n^t = \tilde{d}_n^t + u_n^t$ denotes the load balance with $\tilde{d}_n^t \in \mathbb{R}$ as the net demand and $u_n^t \in \mathbb{R}$ as the control action corresponding to SES.

SES modelling

The agents act "rationally" and can decide on how much power to exchange with the SES and the grid. All the power exchanges are associated with an economic exchange. We model that

- **the agents are billed/compensated if they discharge/charge the SES,**
- **the agents are billed/compensated if they withdraw/inject energy from/into the grid.**

The cost associated with different power exchanges (a.k.a. the electricity bill) can be expressed as

$$b_n^t = \mathtt{max}\{g_n^t c_{\mathrm{ToU}}^t,\, g_n^t c_{\mathrm{FiT}}^t\} - \mathtt{min}\{u_n^t c_{\mathrm{SES}}^{\mathrm{ch},t},\, u_n^t c_{\mathrm{SES}}^{\mathrm{dch},t}\}. \tag{5}$$

In (5), $g_n^t = \tilde{d}_n^t + u_n^t$ denotes the load balance with $\tilde{d}_n^t \in \mathbb{R}$ as the net demand and $u_n^t \in \mathbb{R}$ as the control action corresponding to SES.

. . . cont'd

As the flexibility resource is shared, the constraints are "coupled", i.e.,

- **bounded total charge/discharge**, i.e., $\sum_{n \in \mathcal{N}} u_n \in [\underline{u}, \overline{u}]$,
- **bounded level-of-charge**, i.e., $\mathrm{LoC}^t + \Big[\sum_{n \in \mathcal{N}} u_n\Big] \Delta t \in [\underline{\mathrm{LoC}}, \overline{\mathrm{LoC}}]$,
- **bounded power from the main grid**, i.e., $\sum_{n \in \mathcal{N}} g_n^t \leq \overline{G}$.

Learning goal

to obtain deterministic control policies for each agent, such that $u_n := \mu_n(o_n)$, and schedule the shared storage in a "purely" decentralized manner.

. . . cont'd

As the flexibility resource is shared, the constraints are "coupled", i.e.,

- **bounded total charge/discharge**, i.e., $\sum_{n \in \mathcal{N}} u_n \in [\underline{u}, \overline{u}]$,
- **bounded level-of-charge**, i.e., $\mathrm{LoC}^t + \left[\sum_{n \in \mathcal{N}} u_n \right] \Delta t \in [\underline{\mathrm{LoC}}, \overline{\mathrm{LoC}}]$,
- **bounded power from the main grid**, i.e., $\sum_{n \in \mathcal{N}} g_n^t \leq \overline{G}$.

Learning goal

to obtain deterministic control policies for each agent, such that $u_n := \mu_n(o_n)$, and schedule the shared storage in a "purely" decentralized manner.

. . . cont'd

As the flexibility resource is shared, the constraints are "coupled", i.e.,

- **bounded total charge/discharge**, i.e., $\sum_{n \in \mathcal{N}} u_n \in [\underline{u}, \overline{u}]$,
- **bounded level-of-charge**, i.e., $\mathrm{LoC}^t + \left[\sum_{n \in \mathcal{N}} u_n \right] \Delta t \in [\underline{\mathrm{LoC}}, \overline{\mathrm{LoC}}]$,
- **bounded power from the main grid**, i.e., $\sum_{n \in \mathcal{N}} g_n^t \leq \overline{G}$.

Learning goal

to obtain deterministic control policies for each agent, such that $u_n := \mu_n(o_n)$, and schedule the shared storage in a "purely" decentralized manner.

Reward function

On most occasions the following rewards will be negative, i.e. penalties:

- for **electricity bill**, $\rho_n^b \propto \{\text{tariff differential}, \ \text{magnitude of action}\}$
- for **bounded control action**, $\rho_n^a \propto \{\text{constraint violation}, \ \text{proportion contribution}\}$
- for **bounded level-of-charge**, $\rho_n^{\mathrm{LoC}} \propto \{\text{constraint violation}, \ \text{proportion contribution}\}$
- for **main grid's constraint**, $\rho_n^{\mathrm{NET}} \propto \{\text{constraint violation}, \ \text{proportion contribution}\}$

Reward function

electricity bill

$$\rho_n^b := \begin{cases} a_n(c_{\mathrm{SES}}^{\mathrm{ch}} - c_{\mathrm{ToU}}), & \text{if } a_n > 0, \\ a_n(c_{\mathrm{SES}}^{\mathrm{dch}} - c_{\mathrm{ToU}}), & \text{if } a_n < 0, \end{cases}$$

bounded control action

$$\rho_n^a := \begin{cases} \frac{(\overline{u} - A)[a_n]_+}{\sum_{n\in\mathcal{N}} [a_n]_+}, & \text{if } A > \overline{u}, \\ \frac{(A - \underline{u})[a_n]_-}{\sum_{n\in\mathcal{N}} [a_n]_-}, & \text{if } A < \underline{u} \end{cases}$$

Network constraint at PoCC

$$\rho_n^{\mathrm{NET}} := \begin{cases} \frac{\left(\sum_{n\in\mathcal{N}} g_n - \overline{G}\right)[a_n]_-}{\sum_{n\in\mathcal{N}} [a_n]_-}, & \text{if } \sum_{n\in\mathcal{N}} g_n > \overline{G} \end{cases}$$

bounded level-of-charge

$$\rho_n^{\mathrm{LoC}} := \begin{cases} \frac{(\overline{\mathrm{LoC}} - \mathrm{LoC})[a_n]_+}{\sum_{n\in\mathcal{N}} [a_n]_+}, & \text{if } \mathrm{LoC} > \overline{\mathrm{LoC}}, \\ \frac{(\mathrm{LoC} - \underline{\mathrm{LoC}})[a_n]_-}{\sum_{n\in\mathcal{N}} [a_n]_-}, & \text{if } \mathrm{LoC} < \underline{\mathrm{LoC}}. \end{cases}$$

States, Observations, Actions (tabular setting)

State: The state of the environment at time t is

$$s^t := \left\{ \mathrm{LoC}^t, c^t_{\mathrm{ToU}}, \dots, c^{t+k-1}_{\mathrm{ToU}}, \{\tilde{d}^{t-h}_n\}_{n=1}^N, \dots, \{\tilde{d}^{t-1}_n\}_{n=1}^N \right\}$$

Observation: Instead, the local observation of agent n corresponding to state s^t is

$$o^t_n := \left\{ \mathrm{LoC}^t, c^t_{\mathrm{ToU}} \right\}$$

with **discretized** values.
Action: The control action a^t_n of the agent (equivalent to u^t_n) is discrete as well, chosen from

$$\mathcal{A}_n := \{\underline{u}, \underline{u} + \Delta u, \dots, -\Delta u, 0, \Delta u, \dots, \overline{u} - \Delta u, \overline{u}\}.$$

Consensus$-Q$ (tabular setting)

Algorithm components

- Localization of the Hyper-Q function in (3)

$$Q_n(s, a_n, a_{-n}) \approx Q_n(o_n, a_n). \tag{6}$$

- Distributed linear iterations

$$Q_n^{(k+1)}(\cdot,\cdot) = W_{n,n} Q_n^{(k)}(\cdot,\cdot) + \sum_{n' \in \mathtt{NEI}_n} W_{n,n'} Q_{n'}^{(k)}(\cdot,\cdot), \tag{7}$$

where $\mathtt{NEI}_n$ is the set of neighbours of agent n and $W_{n,n'} \in \mathbb{R}^+$ is a weighing parameter.

Joshi A., M. Tipaldi, and L. Glielmo, "A Consensus Q-Learning approach for Decentralized Control of Shared Energy Storage." *IEEE Control Systems Letters* (2023).

States, Observations, Actions (linear function approximator)

State: The state of the environment at time t is, again,

$$s^t := \{\mathrm{LoC}^t, c_{\mathrm{ToU}}^t, \ldots, c_{\mathrm{ToU}}^{t+k-1}, \{\tilde{d}_n^{t-h}\}_{n=1}^N, \ldots, \{\tilde{d}_n^{t-1}\}_{n=1}^N\},$$

Observation: The corresponding local observation of agent n is richer

$$o_n^t := \{\mathrm{LoC}^t, c_{\mathrm{ToU}}^t, \ldots, c_{\mathrm{ToU}}^{t+k-1}, \tilde{d}_n^{t-h}, \ldots, \tilde{d}_n^{t-1}\}$$

with **continuous** values.
Action: Again, the control action

$$a_n^t \in \mathcal{A}_n := \{\underline{u}, \underline{u} + \Delta u, \ldots, -\Delta u, 0, \Delta u, \ldots, \overline{u} - \Delta u, \overline{u}\}.$$

Consensus$-Q$ (linear function approximator)

Algorithm components

- Localization of the Hyper-Q function in (3)

$$Q_n(s, a_n, a_{-n}) \approx Q_n(o_n, a_n) \approx Q_n\left(o_n, a_n | \theta^{Q_n}\right).$$

- Distributed linear iterations (**federated learning**)

$$\theta_n^{k+1} = \sum_{n' \in \mathtt{NEI}_n} W_{n,n'} \theta_{n'}^k - \alpha_n \frac{\partial f_n}{\partial \theta}\bigg|_{\theta = \theta_n^k},$$

where $\mathtt{NEI}_n$ and $W_{n,n'} \in \mathbb{R}^+$ are the same as above, $\alpha_n \in \mathbb{R}^+$ is the step size and $\frac{\partial f_n}{\partial \theta}\big|_{\theta=\theta_n^k}$ is the subgradient of the objective function locally available to agent n.

Joshi A., M. Tipaldi, and L. Glielmo, "Multi-Agent Reinforcement Learning for Decentralized Control of Shared Battery Energy Storage System in Residential Community," *Sustainable Energy, Grids and Networks* (**revise and resubmit**).

Numerical results

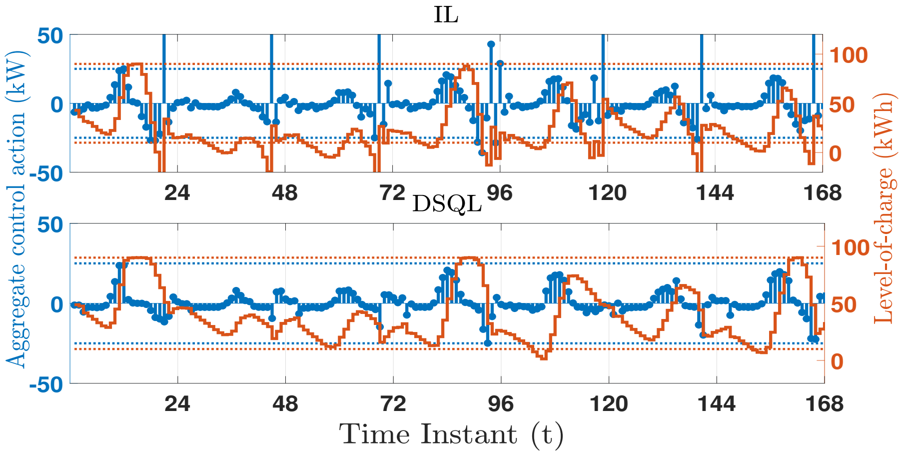
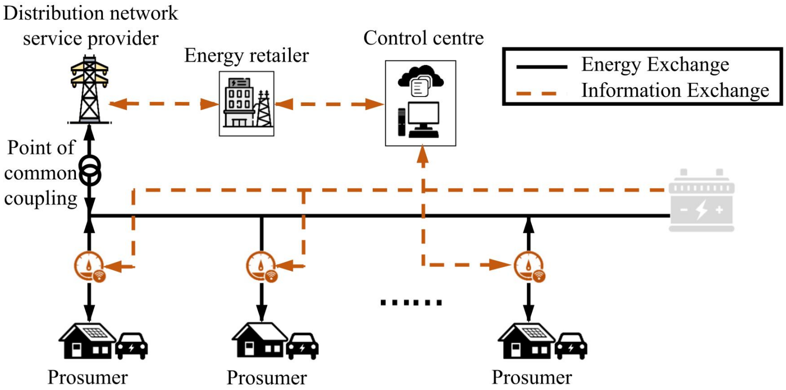


Figure: The aggregate control action and the LoC dynamics over a week when using Independent Learners (IL) and Distributed Subgradient Q-learners (DSQL).

A Case of “Individual” Storage in EVs

Electric Vehicle Coordination (with a belief-based approach)



Conclusions

Conclusions

RL perspective

- Needs historical data
- Sometimes mathematically less rigorous (a problem for safety critical applications)
- Computationally fast
- Time-consuming "reward shaping"

Ongoing initiative

How to integrate the newly developed data-driven techniques with model-based ones summing the advantages of both?

Conclusions

Take away message

The following scenario

- A set of **decision makers jointly coupled** through objective function and/or constraints
- Availability of **historical data** on the exogenous component influencing the decision making

can be translated into a multi-agent coordination problem to tackle through adaptations of model-based and/or the data-driven approaches.

However, as usual. . .

Conclusions

Take away message

The following scenario

- A set of **decision makers jointly coupled** through objective function and/or constraints
- Availability of **historical data** on the exogenous component influencing the decision making

can be translated into a multi-agent coordination problem to tackle through adaptations of model-based and/or the data-driven approaches.

However, as usual. . .

There’s no free lunch!

My Affiliations in Italy

Università degli Studi di Napoli Federico II

Università degli Studi di Napoli
FEDERICO II

Università degli Studi del Sannio di Benevento

UNIVERSITÀ DEGLI STUDI
DEL SANNIO Benevento

Our team for energy and optimization

- Prof. Carmen Del Vecchio (Associate professor at University of Sannio)
- Prof. Davide Liuzza (Assistant professor at University of Sannio)
- Prof. Francesco Lo Ludice (Associate professor at University of Naples Federico II)
- Prof. Pietro De Lellis (Associate professor at University of Naples Federico II)
- Dr. Camilla Ancona (PostDoc at University of Naples Federico II)
- Dr. Luigi Russo (PostDoc at University of Sannio)
- **Amit Joshi (PhD Candidate at University of Sannio)**
- Vishal Kachhad (PhD Student at University of Sannio)
- Emanuele Musicò (PhD Student at University of Naples Federico II)

Thank you!